REMARKS

An Office Action was mailed in the above-captioned application on July 18, 2003. In such Office Action, claims 1-10, 13-18, 20-22, 39, 40 and 42 were pending. Claims 21 and 22 were rejected under 35 U.S.C. § 112, second paragraph. Claims 1-7, 13, 16-18, 20-22, 39, 40 and 42 were rejected under 35 U.S.C. § 102(a) and (e)(2). Claims 1-10, 13-18, 20-22, 39, 40 and 42 were rejected under 35 U.S.C. § 103(a). This Amendment and Remarks is submitted in response to that Office Action.

The specification has been amended to correct a clerical error discovered by Attorneys for the Applicant during review of the application. In particular, the term "lower granularity" has been changed to "finer granularity" in two occurrences in the same paragraph of the specification (¶27; page 9, lines 20-31). This change is supported by its context – measurements of "subspecies of CD4+ T cells" are of "finer granularity" than measurements of "CD4+ T cells [without regard to subspecies]." Likewise, the sentence containing one of the changes refers to bioanalytical instruments that are capable of making "such fine-grained measurements" (emphasis added).

New claims 43-49 have been added to more particularly point out and distinctly claim the present invention. In particular, new claims 43 and 46 depend from claim 1 and recite types of observations. Support for each type of observation can be found at page 2, lines 24-27. New claims 44 and 45 depend from new claim 43 and recite types of clinical classes. Support for each type of clinical class can be found at page 8, lines 12-13. New claims 47-49 depend from claim 1 and recite ranges for n. Support for the ranges of n can be found at page 5, line 6, and page 8, lines 28-29.

No new matter has been added by the new claims or the amendments to the specification. Entry of the foregoing amendments and the following remarks is respectfully requested.

The Present Invention

The pending claims are directed to methods for analyzing broad data sets of biological measurements. Broad data sets are those in which the number of measurements n is larger than the number of observations p, often by several orders of magnitude. As

more fully set forth in the specification, existing statistical and data mining techniques either cannot be applied to broad data sets, or their accuracy is questionable under such conditions. The present invention is a new technique that allows the efficient and accurate extraction of biological markers from broad data sets.

The Rejections Under 35 U.S.C. § 112, Second Paragraph Should Be Withdrawn

Claims 21 and 22 were rejected by the Examiner under 35 U.S.C. § 112, second paragraph because of the phrase "desired computation time." The Examiner states that it is "unclear what is being used to [determine whether?] the computation time is desirable (CPU time or time generated by a computer)." Office Action, ¶8.

Applicant does not believe that "desired computation time" is ambiguous or indefinite. The term has ample description in the present specification. See, e.g., page 14, lines 17-24 ("The user-selected thresholds can be derived based on a desired computation time. For example, the amount of time necessary to perform [a given step] can be determined empirically for a variety of data set sizes. In general, a formula for computation time cannot be determined, because of unknown processor-dependent factors, but the time can be determined empirically. The user can then select a desired computation time, and the required data reduction can be determined from the empirical results."). Thus; the "computation time" for a given task is the amount of time required by the computer to perform the task, however measured; the "desired" computation time is the computation time that is desired.

The term "desired computation time" meets the requirements of 35 U.S.C. § 112, second paragraph, and the rejection should therefore be withdrawn.

The Rejections Under 35 U.S.C. § 102(b) and (e)(2) Should Be Withdrawn

Pending claims 1-7, 13, 16-18, 20-22, 39, 40 and 42 were rejected by the Examiner under 35 USC § 102(b) and (e)(2) as being anticipated by U.S. Patent No. 6,059,724 to Campell et al. ("Campell et al.").

Campell et al. describes a computer-based method for predicting the future health of individuals. According to the method, a set of biological measures is acquired from a large number of patients, each in one of two classes, and the measures are analyzed to locate biological markers capable of distinguishing between the classes. The number of markers to be considered is reduced, and a discriminant analysis is performed on the remaining measures to identify the biological markers. The biological markers can then be used to predict the risk of a person of acquiring a disease corresponding to one of the classes.

The Examiner contends that an example of a training set provided in Campell et al. anticipates claim 1 of the present invention. Office Action, ¶11. In that example, Campell et al. describes a group of 481 patients that collectively made 641 annual visits for evaluation (col. 33, lines 9-14). Thus, not every subject made two annual visits and Campell et al. states that "not all biomarkers were assessed, even when a subject made a visit" (col. 33, lines 11-12). Campell et al. does not disclose how many biomarkers were measured at each visit. However, the Examiner states that other disclosure in Campell et al. suggests the measurement of 200 biomarkers.

Using the number of observations p to represent the number of patients (page 2, line 27), p would be 481 because there were 481 patients from whom samples were taken. See present specification page 2, line 27. To make the data set as broad as possible, Applicant will assume that each of the 200 biomarkers was measured for each patient and that each patient made both annual visits. Making these favorable assumptions, the number of measurements n would be at most 400 (i.e., 200 measurements on two occasions). This is not a broad data set because n (400) is not greater than p (481) as required by claim 1.

The number of observations p can also be used to represent the number of samples (page 2, lines 25-26). Analyzed in this way, however, the data set is even "narrower" – the number of observations p would be no fewer than 641, counting no more samples (observations) than the reference expressly states were obtained and making the favorable assumption that only one sample was taken from each patient at each visit (taking more

than one sample would further narrow the data set). The number of measurements *n* would be 200, assuming that 200 biomarkers were measured for each sample. Despite analyzing the example under conditions most favorable to obtaining a broad data set, *n* (200) is not greater than *p* (641) as required by Claim 1.

Independent Claim 39 requires an even "broader" data set (i.e., $n > 10p$). Campell et al. does not anticipate claim 39 for the same reasons it does not anticipate claim 1.

The Examiner incorrectly states that "200 biomarkers values for each member of the test population, *n*, is greater that of 641 evaluations, *p*." Office Action, ¶11. the Examiner properly finds that the number of measurements *n* is 200, and that the number of observations *p* is 641, just as set forth above. However, claim 1 requires *n* to be greater than *p*, and 200 is not greater than 641.

Similarly, the Examiner incorrectly states that "*n* is 200 biomarker values for each member of the test population 481 and *p* [is] 641 evaluations; therefore $10p$ is less than *n*, as in instant claim 4." Office Action, ¶12. Again, the Examiner properly finds that the number of measurements *n* is 200, and that the number of observations *p* is 641, just as set forth above. However, claim 4 requires *n* to be greater than $10p$, and 200 is not greater than 6410.

The Examiner states that Campell et al. discloses an *n* of 257,932,000 because the reference states that this is the size of the group at risk for heart disease. Office Action, ¶11. However, as used in the present application, *n* refers to the number of measurements made with respect to each patient/sample, not the number of people at risk for developing a disease.

The Examiner also states that Campell et al. discloses obtaining 88 values for direct bilirubin, and the Examiner treats this value as *k*. Office Action, ¶11. However, as used in the present application, *k* does not denote the number of samples/patients for which a given measurement was obtained. Rather, the present application teaches that a biomarker for a particular disease can consist of a number of biological measurements and *k* is used to denote that number (page 9, lines 4-18; page 6, lines 18-25).

Moreover, the method of Campell et al. is directed to a qualitatively different problem from the method of the present claims. In Campell et al., the number of measurements must be reduced sufficiently in order to allow the discriminant analysis to be performed. This is why the number of measurements is smaller than the number of samples. Additionally, in the method of Campell et al. significant domain knowledge is required both to choose the original set of potential markers and later to reduce the set. This domain knowledge does not come from the statistical and data mining techniques themselves – rather, it must be supplied by experts in the field who have knowledge of the particular disease and of the biological factors already known to be important in that disease. This is in sharp contrast to the methods of the present invention which are directed to the problem of searching for markers not previously known to have any correlation with the disease of interest.

The Rejections Under 35 U.S.C. § 103
Should Be Withdrawn

Pending claims 1-10, 13-18, 20-22, 39, 40 and 42 were rejected by the Examiner under 35 USC § 103(a) as being unpatentable over Campell et al. taken with Eisen et al., Proc. Natl. Acad. Sci. USA 95:14863-68 (1998) ("Eisen et al.") in view of U.S. Patent No. 5,871,946 to Lucas et al. ("Lucas et al.").

As described above, Campell et al. does not anticipate the claims of the present application, at least because it does not disclose data sets where $n > p$. The Examiner does not suggest that the secondary references address this deficiency. Rather, the Examiner cites the secondary references for their disclosure with respect to the limitations of correlation based on hierarchical cluster, user-selected correlation threshold and ranking of biological markers. Office Action, ¶23.

Eisen et al. relates to a system for cluster analysis of genome-wide expression data. It is an example of one of the current approaches to the computational analysis of gene expression data that attempt to learn functionally significant classifications of genes. The aim in Eisen et al. is not to locate specific features capable of classifying patients, but rather to cluster different genes into functional classes. Eisen et al. use hierarchical cluster

analysis (HCA) to visualize genes' functional relationships. Based on the cluster trees obtained, a user can hypothesize new gene functional classes. Because of the different purposes and approaches of the references, one of ordinary skill in the art would not be motivated to combine the disclosure of Eisen et al. with that of Campell et al. Moreover, like Campell et al., Eisen et al. lacks a disclosure of broad data sets. Thus, even if combined, Campell et al. and Eisen et al. would not obviate the present claims.

The Examiner states that Lucas et al. discloses a method of studying biological surface markers and ranking cells by the functional activity of the cell markers as in instant claims 14 and 15. Office Action ¶26. However, neither claim 14 nor 15 recite ranking cells – rather, both claims recite ranking biological markers, for example, in dependence on their accuracy of predicting clinical endpoints. The type of ranking disclosed in Lucas et al. relates to classification of cells into various populations on the basis of intracellular enzymatic activity; to "sort cells by type or morphology" (col. 33, lines 35-36). This is entirely different than ranking biological markers, for example, in dependence on their accuracy of predicting clinical endpoints as claimed in the present application.

Moreover, Lucas et al. does not relate to methods of efficiently mining broad data sets. Lucas et al. is directed to a method for determining the activity of an endogenous enzyme in a metabolically active whole cell. The method depends on an "assay reagent" that is able to pass through the cell membrane and come into contact with the targeted enzyme. The assay reagent is specifically designed *a priori* so that cleavage by the target enzyme results in a detectable change in fluorescence. One application of the method is to measure the activity of specific targeted enzymes in diseased cells and compare the result to the activity of those enzymes in a normal cell. The resulting data can then be analyzed by various prior art statistical techniques to determine any relationships between the various enzyme activities.

Because each measurement requires an assay reagent specifically tailored to the targeted enzyme, the method of Lucas et al. is necessarily of limited scope and requires significant domain knowledge. Accordingly, Lucas et al. teaches using a small – not large

– number of measurements. See, e.g., col. 43, lines 52-57 ("The selection of a relatively small number of components for the measurement vector is necessary to simplify the later analysis, reduce the number of physical measurements which must be made and to reduce the effects of spurious noise generated from the measurements and from individual variation among the same population" and then reducing that small number to those "contributing most to distinguishing different disease states" (emphasis added)). Finally, as more fully set forth in the present specification, the prior art statistical methods disclosed in Lucas et al. are inappropriate for analysis of broad data sets. Thus, even if combined, Campell et al. and Lucas et al. would not obviate the present claims.

Conclusion

Applicant believes that the pending claims are in condition for allowance. Thus, Applicant requests that the Examiner reconsider the application and issue a Notice of Allowance in the next Office Action. If it would be helpful to obtain favorable consideration of this case, the Examiner is encouraged to call and discuss this case with the undersigned.

This constitutes a request for any needed extension of time and an authorization to charge all fees therefore to deposit account No. 19-5117, if not otherwise specifically requested. The undersigned hereby authorizes the charge of any fees created by the filing of this document or any deficiency of fees submitted herewith to be charged to deposit account No. 19-5117.

Respectfully submitted,

Date: January 20, 2004

Barry J. Swanson, #33,215
Swanson & Bratschun, L.L.C.
1745 Shea Center Drive, Suite 330
Highlands Ranch, Colorado 80129
Telephone: (303) 268-0066
Facsimile: (303) 268-0065